



HAL
open science

In-situ assessment report of citizen local interactions and self-reporting GEAR cycle 2

Marc Santolini, Camille Masselot, Rathin Jeyaram

► **To cite this version:**

Marc Santolini, Camille Masselot, Rathin Jeyaram. In-situ assessment report of citizen local interactions and self-reporting GEAR cycle 2. European Union's Horizon 2020 research and innovation programme. 2022. hal-04167274

HAL Id: hal-04167274

<https://hal.science/hal-04167274>

Submitted on 20 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Crowd4SDG

Crowd4SDG

Citizen Science for the Sustainable Development Goals

Deliverable 4.4

In-situ assessment report of citizen local interactions and self-reporting GEAR cycle 2

Deliverable identifier: D4.4

Due date: 30/04/2022

Justification for delay: due to exceptional activity as agreed with Project Officer

Document release date: 14/06/2022

Nature: Report

Dissemination Level: Public

Work Package: 4

Lead Beneficiary: UP

Contributing Beneficiaries: UNIGE

Document status: Final

Abstract:

In this report, we present a data-driven approach leveraging digital traces to introduce and monitor metrics pertaining to the diversity, collaboration and activity of teams, and show how these metrics correlate with outcome and process centric criteria of team performance in the context of the GEAR cycle 2. We collected compositional and organizational data from 14 participating teams from different sources: communication data from public Slack channels, surveys, and self-reported activity data from CoSo (Collaborative Sonar), a digital platform for monitoring team collaborations and task management. We then computed four categories of team features, related to team composition (size, background diversity, prior experience with citizen science projects), communication activity (number of posts on Slack, communication with organizers, closeness centrality in the communication network), collaborations (intra-team work interactions, amount and diversity of the advice seeking ties), and teamwork (amount and span of activities performed, regularity of work). Performance was assessed through the final stage achieved by a team, an objective measure of achievement for the team, the quality of the project (relevance to the SDGs, novelty) as judged by experts, and the quality of the process (engagement, deliverables) as judged by the organizing team. Finally, we tested the associations between performance criteria and team features using Pearson's correlations. We found that team composition and activity on Slack are associated with performance measures related to the outcome of the project, while team collaborations and division of labor associates with performance measures related to the elaboration process. We conclude that measures of team composition, activity and collaborations can be leveraged to monitor team performance pertaining to both the project outcome and its elaboration process, highlighting their usefulness in the context of the Crowd4SDG project and similar initiatives leveraging Citizen Science to address the SDGs.

For more information on Crowd4SDG, please check: <http://www.crowd4sdg.eu/>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 872944.



Document history

	Name	Partner	Date
Authored by	Marc Santolini, Camille Masselot, Rathin Jeyaram	UP	April 2022
Revised by	Oguz Mulayim Madina Imaralieva	CSIC UNITAR	26/04/2022 03/05/2022
Edited by	Marc Santolini, Camille Masselot, Rathin Jeyaram	UP	06/05/2022
Reviewed by	Jose Luis Fernandez-Marquez	UNIGE	25/05/2022
Approved by	François Grey Jose Luis Fernandez-Marquez	UNIGE	14/06/2022

Table of Contents

Document history	2
Project Partners	4
Crowd4SDG in brief	5
Grant Agreement description of the deliverable	6
1. Introduction	8
1.1. Context	8
1.2. Changes made compared to the GEAR cycle 1	8
2. Research questions	10
3. Methods	11
3.1. Data collection	11
3.1.1. Slack data	11
3.1.2. CoSo data and surveys	11
3.2. Team performance	13
3.3. Data analysis	14
3.4. Summary description of data frame in Zenodo (deliverable 4.2)	15
4. Results	16
4.1. Descriptive statistics of the diversity, activity and collaborations among teams	16
4.1.1. Registration Form analysis: team profile and diversity	16
4.1.2. Slack analysis: team activity	18
4.1.3. CoSo analysis: teamwork and collaborations	22
4.2. Team performance analysis	24
5. Discussion and perspectives	27
6. Conclusion	28
7. References	29
Annex 1: list of abbreviations	30
Annex 2: Surveys	31
Registration Form	31
Evaluate weekly 1, 2, 3 and Accelerate	31
Final form - Evaluate	33
Accelerate initial Form and bloc 1	33

Project Partners

	Partner name	Acronym	Country
1 (COO)	Université de Genève	UNIGE	CH
2	European Organization for Nuclear Research	CERN	CH
3	Agencia Estatal Consejo Superior de Investigaciones Científicas	CSIC	ES
4	Politecnico di Milano	POLIMI	IT
5	United Nations Institute for Training and Research	UNITAR	CH
6	Université de Paris	UP	FR



Crowd4SDG in brief

The 17 Sustainable Development Goals (SDGs), launched by the UN in 2015, are underpinned by 169 concrete targets and [231 unique measurable indicators](#). Some of these indicators initially had no established measurement methodology. For others, many countries do not have the data collection capacity. Measuring progress towards the SDGs is thus a challenge for most national statistical offices.

The goal of the Crowd4SDG project is to research the extent to which Citizen Science (CS) can provide an essential source of non-traditional data for tracking progress towards the SDGs, as well as the ability of CS to generate social innovations that enable such progress. Based on shared expertise in crowdsourcing for disaster response, the transdisciplinary Crowd4SDG consortium of six partners is focusing on SDG 13, Climate Action, to explore new ways of applying CS for monitoring the impacts of extreme climate events and strengthening the resilience of communities to climate related disasters.

To achieve this goal, Crowd4SDG is initiating research on the applications of artificial intelligence and machine learning to enhance CS and explore the use of social media and other non-traditional data sources for more effective monitoring of SDGs by citizens. Crowd4SDG is using direct channels through consortium partner UNITAR to provide National Statistical Offices (NSOs) with recommendations on best practices for generating and exploiting CS data for tracking the SDGs.

To this end, Crowd4SDG rigorously assesses the quality of the scientific knowledge and usefulness of practical innovations occurring when teams develop new CS projects focusing on climate action. This occurs through three annual challenge based innovation events, involving online and in-person coaching. A wide range of stakeholders, from the UN, governments, the private sector, NGOs, academia, innovation incubators and maker spaces are involved in advising the project and exploiting the scientific knowledge and technical innovations that it generates.

Crowd4SDG has six work packages. Besides Project Management (UNIGE) and Dissemination & Outreach (CERN), the project features work packages on: Enhancing CS Tools (CSIC, POLIMI) with AI and social media analysis features, to improve data quality and deliberation processes in CS; New Metrics for CS (UP), to track and improve innovation in CS project coaching events; Impact Assessment of CS (UNITAR) with a focus on the requirements of NSOs as end-users of CS data for SDG monitoring. At the core of the project is Project Deployment (UNIGE) based on a novel innovation cycle called GEAR (Gather, Evaluate, Accelerate, Refine), which runs once a year.

The GEAR cycles involve online selection and coaching of citizen-generated ideas for climate action, using the UNIGE Open Seventeen Challenge (O17). The most promising projects are accelerated during a two-week in-person Challenge-Based Innovation (CBI) course. Top projects receive further support at annual SDG conferences hosted at partner sites. GEAR cycles focus on specific aspects of Climate Action connected with other SDGs like Gender Equality.

Grant Agreement description of the deliverable

The focus of Work Package 4 (WP4), led by the University of Paris, is to conduct research on Citizen Science. This encompasses the establishment of methods and the collection of data to inform the development of effective, high-quality citizen science projects. To that aim, this work package develops metrics and statistical models in order to assess the many-faceted outcomes of the citizen science projects developed within the Crowd4SDG consortium.

By leveraging the digital traces from online tools that document project progress and citizen engagement, it is possible to quantitatively monitor and analyse i) the activity of teams working within the GEAR collaborative framework, and ii) the activity and engagement patterns of citizen science participants. As such, we propose to frame the participants' behaviour, their interactions and engagement with tools used during the GEAR cycles.

We highlight in bold what this deliverable addresses among the tasks of WP4.

T4.2: Measuring analytics of citizen collaborations using new metrics/descriptors on digital traces (UP, UNIGE)

The epistemological analysis performed in Task 4.1 provides the conceptual foundation for specific new metrics and descriptors for Citizen Science. While individual learning progress can be monitored in a straightforward manner by the increase in relative level of expertise on specific keywords/topics (see, for example, the iLearn project at CRI), it remains unclear how such insights can be generalised to assess overall learning progress achieved in specific challenges/projects and the entire scope of projects initiated within the scope of Crowd4SDG. For this reason, this Task will **develop and implement new metrics and descriptors that not only measure productivity and output, but also assess the overall learning and research dynamics as well as the diversity, originality, relevance, robustness, and adaptiveness of the knowledge produced in the context of each individual citizen-innovation team as well as the entire group of citizen-participants across all CS projects within Crowd4SDG.** The work involves the development of specific algorithms based on a fundamental discussion of the kind of knowledge Citizen Science is expected to produce and whether/how it differs from knowledge produced by conventional scientific approaches (see Task 4.1). In the proposed project, participants will have access to and assess—in a distributed, large-scale manner—the local impact of the Crowd4SDG program. By monitoring the activity patterns of participants when they use the Citizen Science Solution Kit (examples: project documentations of SDG in Progress or the community management solutions of CrowdBuilder) to conduct real time analyses of the collaboration within the teams, we can frame the participants' behaviour, their interactions and engagement within the O17 Challenges that form part of the project.

These digital traces will be used to explore the previously defined metrics of citizen science quality. The work in this task will build upon previous work by the partners. Using data from the various tools offered by the CS Solution Kit (see section 1.3.4.2), namely SDG in Progress and Pybossa, we will, in a similar manner, quantify team diversity (skills and backgrounds), dynamics (bursts of activity), organisational structure (core-periphery network structure and leadership dynamics), as well as the influence of physical meetups on team activity.

T4.3: In-situ assessment of citizen local interactions and self-reporting (UPD, UNIGE)

In this Task a self-report smartphone application will be developed in order to be able to establish deeper insights into collaboration dynamics. The app will be able to assess social

proximity through smartphone sensors (Bluetooth, Wi-Fi and GPS) and trigger context-based notifications. We intend to use this app to provide fine-grained measurements of citizen collaborative dynamics, both in terms of social network dynamics and in terms of learning experience through the self-reporting contextualisation it offers. The application will be based on an open-source platform developed by researchers at the Child Mind Institute in New York in the context of a large-scale mental health study. Their platform, Mindlogger, is a general-purpose open source data collection platform that can be used by anyone to administer surveys, quizzes or different types of tasks. Proximity-based notifications will be implemented to trigger "Ecological Momentary Assessments" asking individuals to document the type of interaction they are currently part of, so that we obtain an augmented understanding of the social context. **Lastly, we will provide reports that provide insights into network structure and summary statistics.** The reports will be in the form of a dashboard tool based on open source software MITeams. The results of this task will be used for assessing participant interactions during the different phases of the GEAR methodology cycle and thus feed back into WP3. **The self-reporting phone app will be used for the in-person events to be carried out in the Accelerate Phase carried out (see Task 3.2).**

1. Introduction

1.1. Context

WP4 aims to develop and monitor new metrics and develop statistical models of team engagement and collaboration that contribute to the many-faceted outcomes of the citizen science projects developed within the Crowd4SDG consortium over the 3-years course of the project.

The WP4 has two specific objectives:

- Develop **new standardised metrics and descriptors** for measuring the diversity, originality, effectiveness, sustainability/robustness and adaptation/appropriateness of solutions and insights gained from Citizen Science projects.
- Implement the metrics and descriptors as tools to **analyse the digitised records** of Citizen Science collaborations and the solutions and insights they produce.

By doing so, WP4 contributes to the following specific objectives of the Crowd4SDG project:

- SO1.2: “Create CS projects and study the mechanisms that lead to improved citizen science skills and high-quality scientific outcomes”, and
- SO2.2 “Produce economic and social outputs relevant to achieving SDGs through challenge-based CS events, with a special focus on climate change resilience”

1.2. Changes made compared to the GEAR cycle 1

In the GEAR cycle 1 report, we presented a preliminary data-driven approach to describe participant demographic profiles, team diversity, activity and interaction dynamics. We introduced the CoSo platform for collecting self-reported data on collaborations and task allocation. CoSo was developed to overcome the setback introduced by the shift to a fully online program due to the COVID pandemic, and the subsequent impossibility to implement a sensor-based contact-tracing app to monitor *in-situ* team interactions using smartphone sensors (Bluetooth, Wi-Fi and GPS). Our findings served as a basis for i) exhibiting the potential of using digital traces from online tools to derive measures related to team process, ii) highlighting perspectives for measuring metrics and descriptors in the next GEAR cycles. We also showed the importance of building a unified framework for collecting a homogenous dataset across all phases of the GEAR cycle, as well as the need to foster participant engagement on a few select platforms.

During the second year of the Crowd4SDG project, we worked with consortium partners to unify the tools used by the participants in order to access consistent, homogenous data. We leveraged the use of email addresses as a single user identification system for all tools and collected self-reported data on CoSo and Google Forms. We collected comprehensive profile information about participants along with temporal insights on their collaborative activities. We also included internal evaluation questions to avoid extensive manual curation from the organising team and ensure analysis-ready data.

Noticeable differences between the GEAR cycle 1 and 2 have to be taken into account when looking at team diversity, activity, and collaboration. For this GEAR cycle:

- A teaming algorithm was used to form teams among participants who joined as individuals (see Deliverable 3.4)

- The Accelerate phase changed format to switch from a 2-week intensive workshop with daily sessions, to a 1-day per week workshop for 5 weeks that were spread between December 2021 and February 2022.

2. Research questions

In this report, we provide an analysis to answer the following questions related to the GEAR cycle 2:

- What are the relevant metrics that can be monitored to assess the quality of citizen science projects?
- To what extent can digital traces and self-report data capture the diversity, collaboration and activity of participants?
- How do diversity, collaboration and activity metrics associate with various criteria of team performance during the GEAR cycle 2?

3. Methods

In this section, we describe how we collected data from Slack and surveys to capture the diversity, collaboration and activity of participants, and the methods we used for testing the correlation between these metrics and the success of teams.

3.1. Data collection

3.1.1. Slack data

A Slack workspace was used by the teams during the different phases of the GEAR cycle as a means to communicate with other teams and with the organising team. This data source was shown to provide insights on team interactions during GEAR Cycle 1, and we decided to build a pipeline to extract and analyse the traces. Different Slack channels were used for each phase. Here, we focus on the “Evaluate” phase, consisting of coaching events and providing the largest team base (14 teams) since it contains the largest amount of data to enable us to identify early predictors for both final success and dropout.

The extracted data consists of:

- User Metadata: This includes data from the workspace members profiles, such as the Slack ID, user name, profile description, timezone, status (admin, bot or others).
- Channel Metadata: This includes the Channel ID, description, creators, members, attributes (is private, is shared) and pinned messages.
- Messages: Each message has an ID, timestamp, sender, text, reactions (users who reacted to the message with an emoticon), and a reference ID to a parent message if it is a part of a thread.

3.1.2. CoSo data and surveys

While Slack allows us to gather information on communication between participating teams and the organising team, it does not encompass the totality of direct information about team activity (division of labour) or collaborations (both formal and informal, such as advice seeking). As was evoked in our previous report for GEAR 1, here we leveraged the newly developed CoSo platform (Tackx et al., 2021) to disseminate a total of 12 surveys on a weekly basis through Slack and during weekly sessions, with a single sign-in system allowing for seamless data integration (Figure 1). A summary of survey description is provided in Table 1. The full surveys were filled online on <https://igem-ties.info/crowd4sdg/> and can be found in Annex 2.



Who are we ?

At Crowd4SDG, we are using surveys for two purposes: improving your experience with the programme, and exploring how activity, interactions, and diversity of teams impact their success. Thank you for helping us to do so! You can find more about it in this [information form](#).

To access the survey, please login using your email and select Participants as a team name.

Team

Select your team

Email (the mail email you gave in the registration form)

Enter your email

Submit

How helpful did you find the session on [name of the session] ?

- Did not attend & did not watch the Zoom recording
- Very unhelpful
- Unhelpful
- Neither unhelpful nor helpful
- Helpful
- Very helpful

Which task(s) did you perform since the previous block?

- Brainstorming / Ideation
- Planning tasks
- Developing protocols / methods
- Reading papers / documents
- Attending meetups with other teams
- Collaborating with other teams
- Interpreting results
- Preparing other documents/materials to share outside the team
- Preparing graphics / videos
- Writing / preparing a pitch
- Project administration
- Software development
- Hardware development
- Data analysis / modeling / visualization
- Meeting with people affected by the problem you are trying to solve
- Meeting with actors of the field of your solution
- Team Building activities
- Documentation of your project
- Use of CS project builder
- Use of CS logger
- Use of Decidim
- Use of Visual Cit
- Use of SDG in progress

Who did you seek advice from during the previous block?

Select the persons by name from the list below. You may need to scroll down to see all of them.

Search the participants

Who did you work with since the previous block?

Select collaborators within your team or from other teams. You may need to scroll down to see all of them.

Search the participants

Figure 1: Screenshots of the CoSo platform and the CoSo surveys disseminated weekly

On CoSo, participants were asked to answer the following questions at the end of each week:

- Who did you collaborate with last week?
- Who did you seek advice from last week?
- How helpful did you find [name of the session]? This question was used for evaluation of the GEAR cycle which results are presented in the D3.4.

In parallel, consortium partners joined forces to co-design a registration survey and a final survey for each phase. These were collected via Google Form. They included profile information and all internal evaluation questions. Reminders were sent on Slack for the weekly surveys and via email for the Final ones. Results are presented in the [D3.4 “GEAR report cycle 2”](#)

The surveys and research questions received the ethical approval of the Inserm committee attached to the University of Paris team (IRB00003888), in charge of collecting the data. Participants gave their consent to the collection of data as they registered to the Evaluate phase (See Annex 2. Registration Form). Data was pseudo-anonymized by our team before analysis. With the exception of the team “Flood Rangers” who answered only one CoSo survey, at least one member of each team answered surveys each week. Therefore, we have a representation of teams close to 100%.

Survey	Data collection platform	Respondents (% of total participants)
Evaluate registration Form	Google Form	38 (100%)
Evaluate weekly 1	CoSo	26 (68%)
Evaluate weekly 2	CoSo	26 (68%)
Evaluate weekly 3	CoSo	20 (53%)
Evaluate weekly 4	CoSo	22 (58%)
Evaluate Final Form	Google Form	22 (58%)
Accelerate initial Form and bloc 1	CoSo	11 (79%)
Accelerate bloc 2	CoSo	12 (86%)
Accelerate bloc 3	CoSo	11 (79%)
Accelerate bloc 4	CoSo	6 (43%)
Accelerate bloc 5	CoSo	7 (50%)
Accelerate Final Form	Google Form	11 (79%)

Table 1. List of surveys of the GEAR cycle 2.

3.2. Team performance

For this report, we consider that performance is the capacity of a team to reach the next stage of the GEAR cycle. We base our analysis on the scores and grids provided by the organisation team to select the teams for the Evaluate phase (Annex 2), allowing us to build early predictors of eventual team performance.

At the end of each phase, experts composing a jury scored each team from 0 to 5 on 5 criteria. We indicate the weight of each score between squared brackets, summing to a total jury score of 50.

- Novelty: Is the pitch based on a new idea or concept or using existing concepts in a new context? [10]
- Relevance: Is the solution proposed relevant to the challenge or potentially impactful? [10]
- Feasibility: Is the project implementable with reasonable time and effort from the team? [10]
- Crowdsourcing: Is there an effective crowdsourcing component? [10]
- Overall: How would you rate this team's overall presentation skills during this pitch? [10]

Between the Evaluate phase and the Accelerate phase, additional criteria presented below were used by the organisation team. We indicate the weight of each score between squared brackets, summing to a jury score of 40

- Appropriateness of Methodology [5]
- Weekly Evaluation [10]
- Use of Toolkit [5]
- Data Collection and NSO [5]
- Commitment [5]
- Attendance [5]
- Deliverables [5]

The final score accounted for 60% of the jury score and 40% of the organisation team score:

$$\text{Final Score} = \text{jury score} * (60/50) + \text{organisation team score.}$$

3.3. Data analysis

The CoSo and Slack data were analysed using the R software. Codes and anonymised data can be found in the open source [Github repository](#).

Using the available Slack data, we used the number of posts and number of reactions of a user as a marker of individual engagement, or team engagement when aggregated over team members. Furthermore, we built social interaction networks where a user is linked to another user if he/she mentions him/her, with a weight corresponding to the number of mentions. This allows to represent the flow of information characterising this phase, in particular highlighting the interactions with the organisation team.

CoSo was used to collect interactions between participants. This information was aggregated over all timepoints collected, yielding weighted interaction networks where edge weights correspond to the number of times the interaction was reported. Network centrality measures were computed using the igraph library in R.

We explored different metrics that correlate positively or negatively to the success of teams. We tested the association between performance (dependent variables) and team features (independent variables) using Pearson's correlations. The p value for the correlation is calculated by computing the *t statistic* (*cor.test* function in R), with the null hypothesis that the correlation between the dependent and independent variable is 0. The level of significance is set at $p = 0.1$.

3.4. Summary description of data frame in Zenodo (deliverable 4.2)

A description of metrics and descriptors presented below and the associated database can be found in the [Deliverable D4.2](#) and on Zenodo: <https://doi.org/10.5281/zenodo.6583538>.

4. Results

In this section we first provide descriptive statistics of the collected Slack and CoSo data, and then introduce derived features that are associated with performance.

4.1. Descriptive statistics of the diversity, activity and collaborations among teams

4.1.1. Registration Form analysis: team profile and diversity

We first provide a summary of the team profile and diversity, which is expanded in D3.4 ([link](#)).

Overall, the background of participants and their age distribution showed a higher diversity within teams than the other measures such as nationality, level of education, occupation, gender, and prior experience.

Even though 17 nationalities were represented during the #Open17ClimateGenderchallenge, which is higher than during the GEAR cycle 1 (Figure 2a), only four teams had members of different nationalities. Most teams (12/14) had members whose current or highest level of education was university (Figure 2b). There was a high level of homogeneity regarding level of education within teams. Indeed, only “Water Warrior” and “Climate Gender Justice” had diverse ones. While all teams were composed of students, it is to note that some of them cumulated employment and entrepreneur positions (Figure 2c).

On the other hand, high interdisciplinary backgrounds were observed within teams (Figure 3). Not only were they formed of members from diverse background (Figure 3a), spanning fields from psychology to agriculture (“Womer”) or from mathematics to social sciences and business (“Andapé Institute”), but members also had a diverse set of skills at the individual level (Figure 3b).



Figure 2: Sankey diagrams of teams (left) and various features of the participants on the right: **a.** their country of origin, **b.** current or highest level of study, and **c.** occupation.

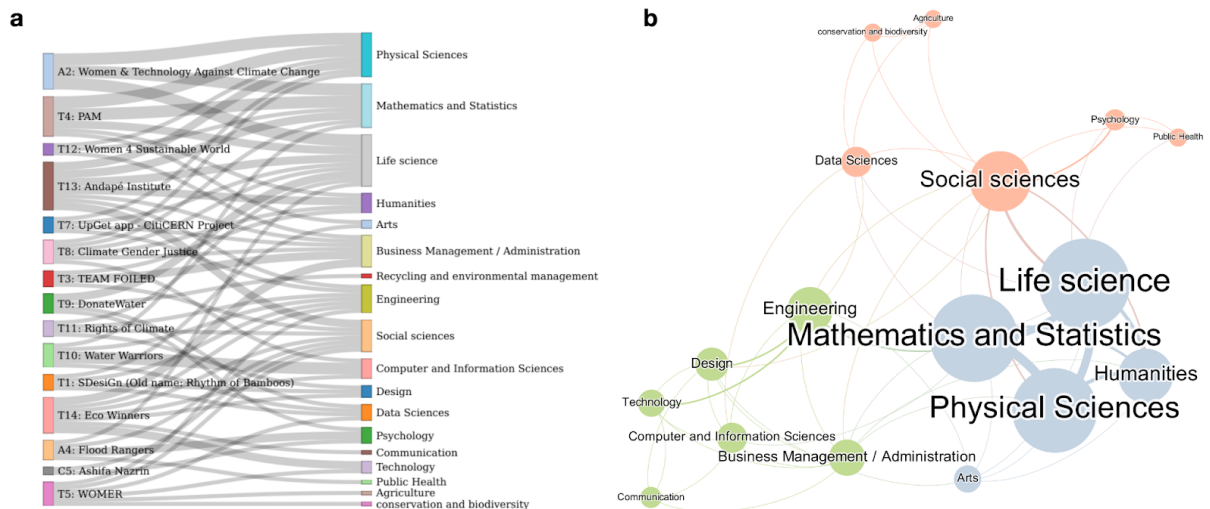


Figure 3. **a.** Sankey diagrams of teams (left) and participant disciplinary backgrounds (right). **b.** Network of academic backgrounds linked by the number of times they are co-reported by a participant (Evaluate phase). Colors correspond to communities of shared skills, as determined by the modularity algorithm.

Ensuring gender balance and enabling female representation was core of the #Open17ClimateGender challenge. We find that all teams, with the exception of the team “Water Warrior” and “Eco winner”, showed a gender balance geared towards more female representation (Figure 4a) with at least 50% of their members identifying themselves as such (Figure 4b). Of 38 participants, 68.4% were between 15 years and 25 years old. The mean age across teams range from 16 years old, to 32 years old, with a high intra-team age span (Figure 4c).

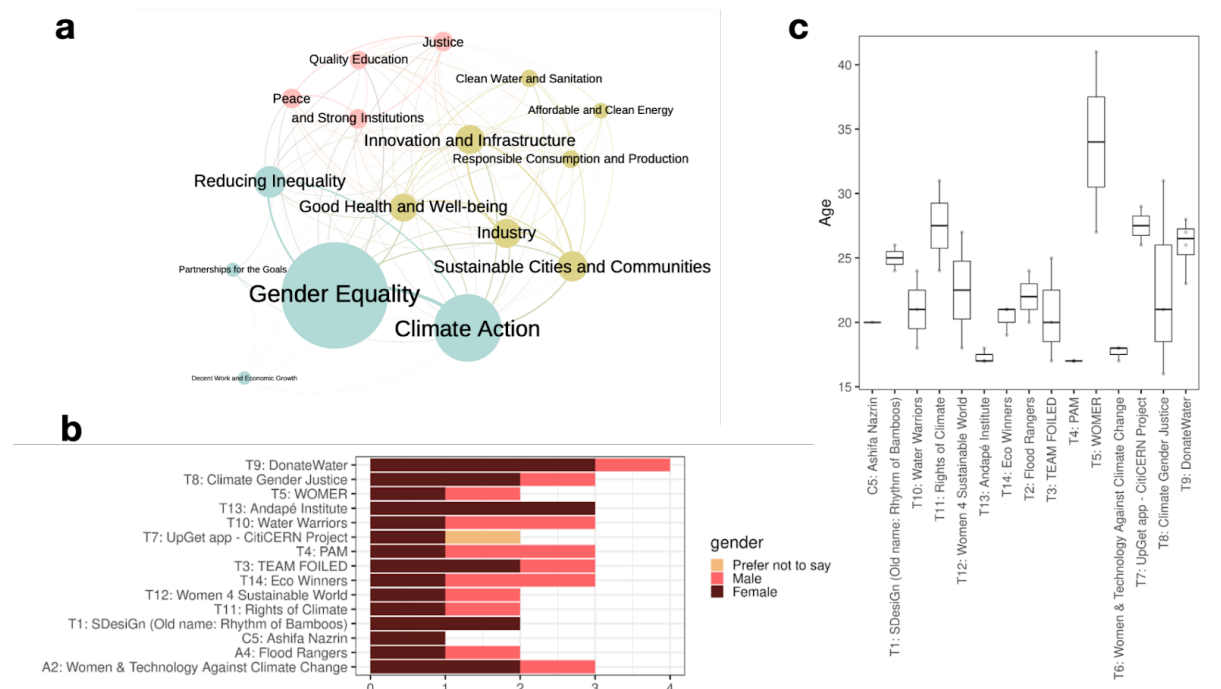


Figure 4: **a.** Network of co-occurrence of SDGs co-reported among GEAR 2 projects. Edge weights denote the number of projects that mention both SDGs. The theme of the cycle was Climate Action and Gender Equality, which appears as the two main hubs of the network. **b.** Gender repartition of team during the Evaluate phase. **c.** Age distribution of participants to the Evaluate phase.

Finally we note that some participants had prior experience with citizen science: to the question “Have you participated in data projects or contributed as citizen scientist to data production before?”, 10 people out of 38 answered yes.

4.1.2. Slack analysis: team activity

Using time-stamped Slack posts, here we describe the activity of the community (including participants, organisation team, and mentors) during the Evaluate phase.

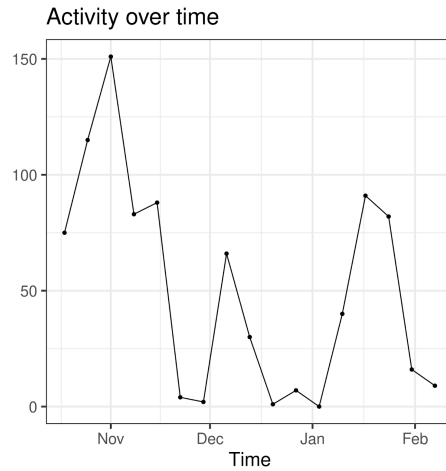


Figure 5: Total number of Slack posts per week in public channels.

We observe bursts of activity on Slack during the last week Evaluate phase (November peak), and the first (December peak) and last weeks (January peak) of the Accelerate phase (Figure 5), highlighting deadline effects. Overall, people were less active during this GEAR cycle compared to GEAR 1. In addition, the drop of posting in public channels during the Accelerate can be attributed to the fact that Slack channels dedicated to this phase were made private, and are not collected here.

Team “Donate Water”, which is one of the two finalists, was the most active in their channel “team-donate-water-project” during the Evaluate phase. With 400 posts in their channel, they were the most active team of all GEAR cycles considered, four times more active than the most active team of the GEAR1.

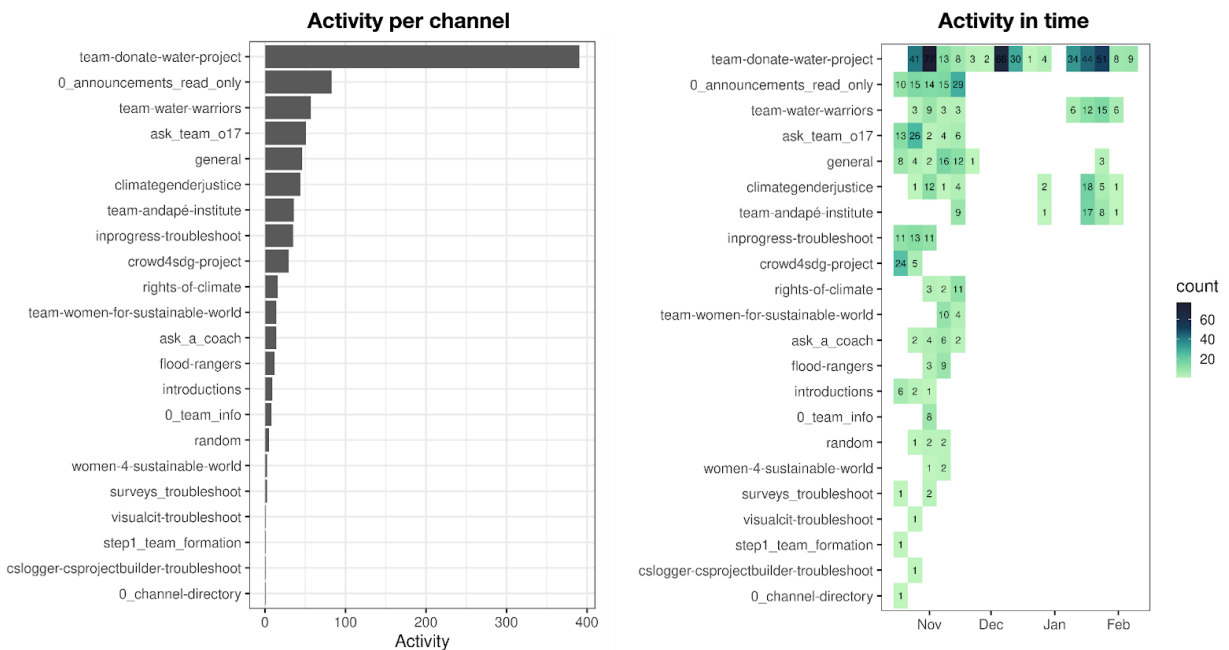


Figure 6. Left: Barplot showing the number of posts per channel. Right: Heatmap showing the temporal distribution of posts across channels. For each channel, we compute the number of messages posted a given week (number of posts that week divided by total number of posts in that channel). Colours range from white (low) to green (high).

Beyond individual post information, Slack offers the opportunity to identify interactions between individuals, through their reaction to posts of other members. This allows us to gather insights on the quantity of interactions at the intra-team, inter-team, as well as mentors/organizers team levels. To examine the interaction dynamics, we use a network approach, with nodes representing individuals (designated by their role or the name of their team) or teams (aggregated data across team members) linked by their interactions. In the following figures, node size and link thickness are proportional to the number of interactions (at the individual or pair level). The network approach allows us to gather insights on i) who are the most central individuals in the information flow and ii) how much horizontal communication across the community is observed.

In order to reconstruct interaction networks from Slack data, we need to define a unit of interaction. This is challenging, since text-based interactions differ from traditional face-to-face interactions or friendship networks used in social network studies (Poquet et al., 2020). Here, we combined two approaches. First, we reconstructed a “direct interaction network” utilising user mentions. These correspond to explicit mentions of a user by another user, using an “@” tag, and represent specific calls for help/action (as such mentions notify another user). In addition, we reconstructed a (denser) network of reactions to posts using emoji reactions. These represent more discreet interactions that signify quick approval (similar to nodding to signal understanding). The aggregation of both networks was used for subsequent analyses.

We show both networks aggregated at the team level in Figure 7. Links between two nodes indicate the number of interactions between team members of two teams, and self-loops indicate interactions within teams. Links are directed from sender to receiver. We find that the mention network (Fig 7) has a strong “star-shaped” centred around the organisation team. This indicates the centrality of the organisation team, with few horizontal interactions between teams. The interactions are bidirectional, meaning that teams both ask for and receive help from the coaching team. When looking at the reaction network, we observe that while organisers are still central, the network is more distributed, indicating quick approval between messages from different teams, without further discussions. Overall, most of the interactions go from teams to organisers, suggesting that the communication is top-down—the organisers provide information in a post and team members approve it.

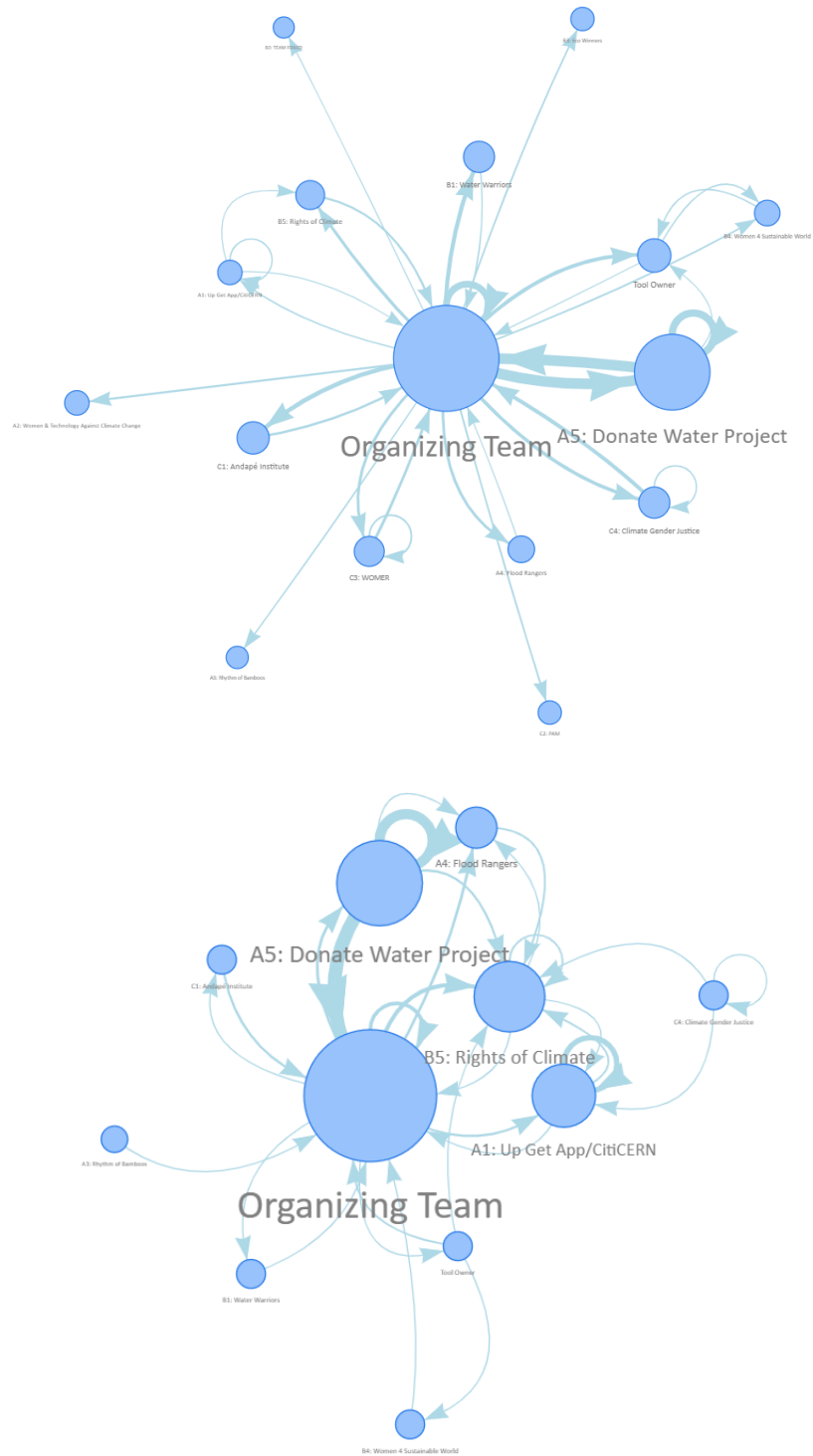


Figure 7: Communication networks from Slack during the Evaluation and Accelerate phases. Nodes represent aggregated data at the team level. **a** Mention network. **b** Reaction network. Teams are linked by weighted edges quantifying the number of times an individual from one team reacts to a post from another team.

4.1.3. CoSo analysis: teamwork and collaborations

In addition to Slack data, we used CoSo to measure the activities that teams performed, and the structure of the collaboration network they constituted over the cycle.

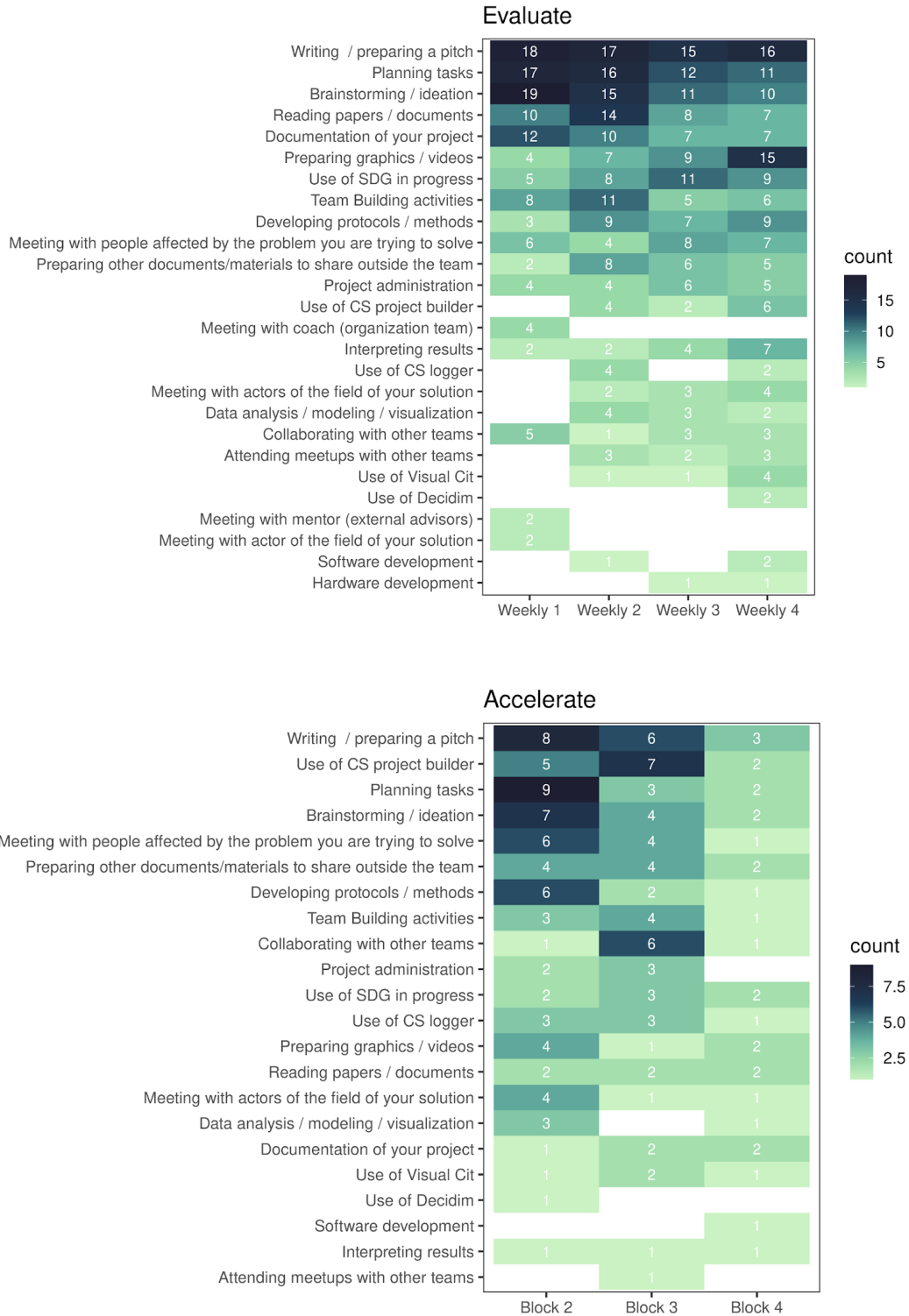


Figure 8: Heatmap of activity of participants over time during the Evaluate phase (left) and the Accelerate phase (right). The count is the number of participants declaring performing a certain activity at the end of each week.

During the Evaluate phase, the realisation of tasks followed closely the agenda of the Open17 program (figure 8). During the first and second week, participants focused on writing, planning and brainstorming while the use of tools remained low. During the fourth and last week of Evaluate, participants spent more time preparing videos and pitches in preparation of the final jury assessment. The use of tools slowly progressed from the second to the fourth week of Evaluate but remained low compared to other types of activities. Participants mainly used [InnProgress](#) (previously SDG in Progress) during the third week to document their project.

During the Accelerate phase, the use of [CS Project Builder](#) ranked second in the most performed task, just after writing and preparing a pitch which shows a progression towards more prototyping of their project. Similarly, meeting with people affected by the problem and preparing material to disseminate their project were more often performed than during Evaluate, confirming the maturation of the projects between these two phases. However, the use of other tools such as [VisualCit](#) and [Decidim4CS](#), meeting with actors of the field and attending meetings with other teams remained the least performed activities.

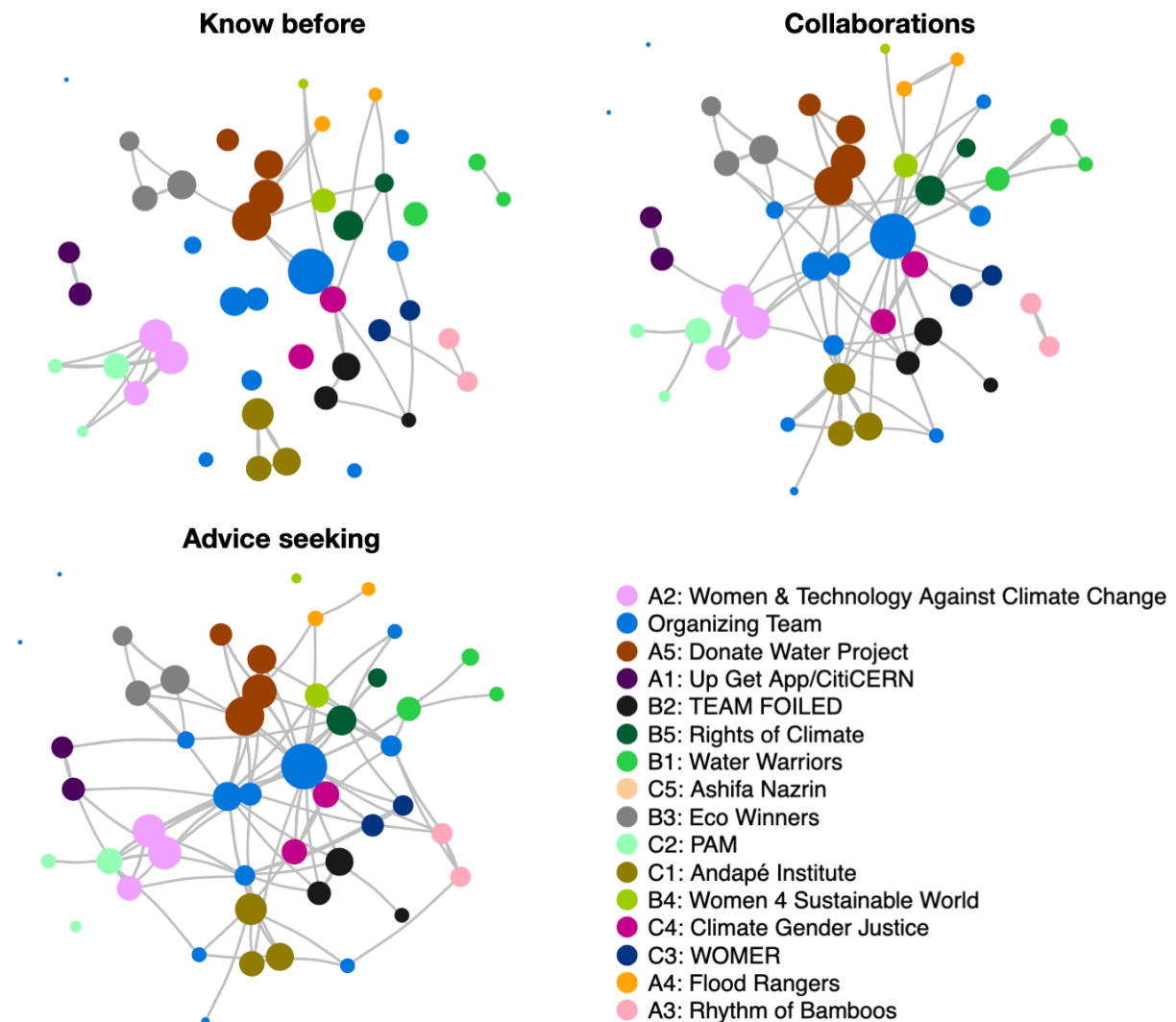


Figure 9: Participant networks constructed from self-report data from CoSo, using prompts: “Which of these people did you know personally before?”, “Who did you work with last week?”, and “Who did you

seek advice from last week?”. The size of the node is proportional to the degree (number of links) in the overall network.

Finally, CoSo was used to collect self-report data on interactions between participants, allowing us to reconstruct the structure of the informal (advice seeking and prior knowledge) as well as formal (work collaborations) networks of the participants (figure 9). We observe that while some teams are constituted by participants who knew each other before, there are also a limited number of inter-team prior ties, showing that the challenge was propagated within existing networks and mobilised to some extent the sphere of influence of participants. We find that the formal work collaborations and advice-seeking networks show similar patterns, with a high intra-team density (team cohesiveness) as well as a large number of connections to the organising team members, which appear as central connectors.

4.2. Team performance analysis

To quantify team success, we used the scores that teams obtained in their assessment by the jury and the Crowd4SDG team. We also used the total score which was used to select the top 5 teams that moved to the Accelerate phase. Finally, we used the furthest stage of advancement per team as another outcome measure (1 - evaluate , 2 - accelerate, 3 - refine), allowing us to measure if early signs measured during the evaluate phase are indicative of future performance (Table 2).

Data can be found in the associated database of the [Deliverable D4.2](#).

Team name	Furthest phase reached
T5: WOMER	Refine
T9: DonateWater	Refine
T14: Eco Winners	Accelerate
T13: Andapé Institute	Accelerate
T10: Water Warriors	Accelerate
T12: Women 4 Sustainable World	Evaluate
T1: SDesiGn (Old name: Rhythm of Bamboos)	Evaluate
T6: Women & Technology Against Climate Change	Evaluate
T8: Climate Gender Justice	Evaluate
T7: UpGet app - CitiCERN Project	Evaluate
T4: PAM	Evaluate
T3: TEAM FOILED	Evaluate
T11: Rights of Climate	Evaluate
T2: Flood Rangers	Evaluate

Table 2: Furthest stage reached by teams.

Given the small number of teams from which we can compute an association with performance (20 data points), we use a correlation analysis with a soft significance threshold at $p = 0.1$. As such, the associations denote trends that will need to be validated by replicating this analysis in the GEAR 3 data. We present the results in Figure 10 below.

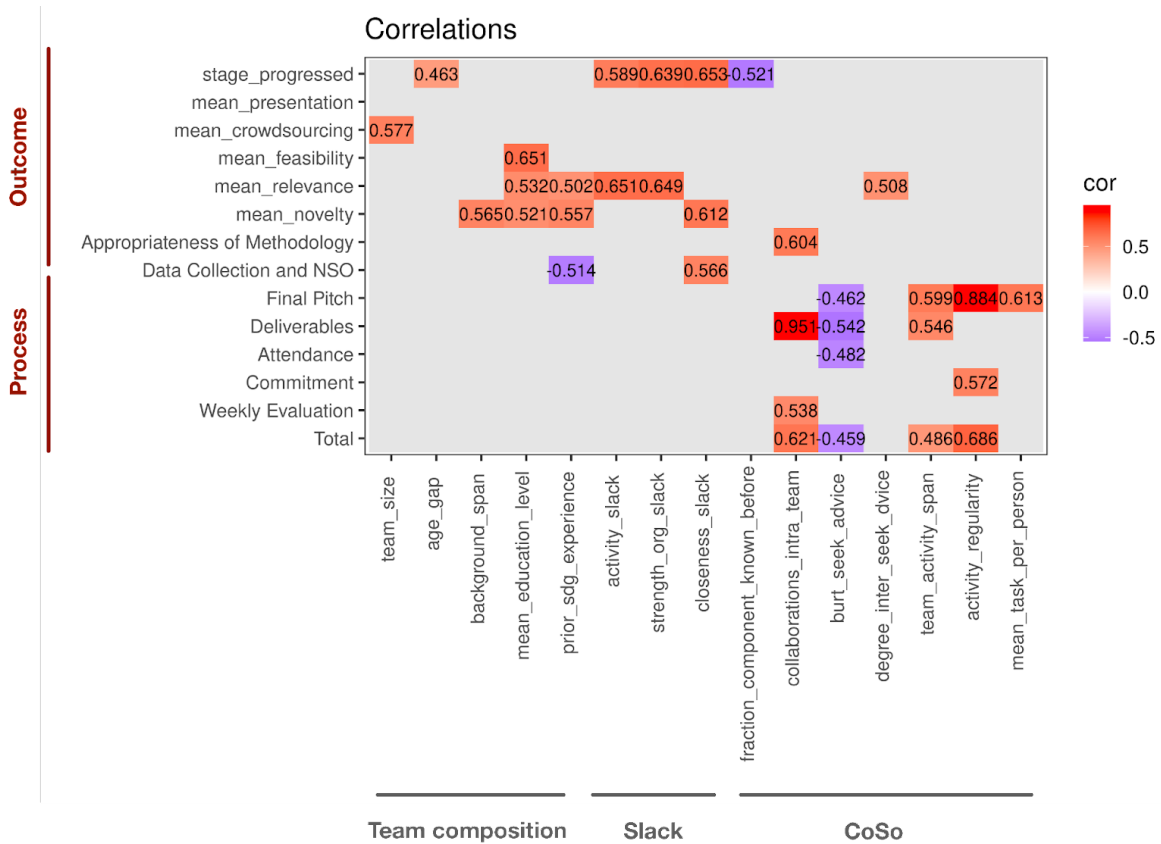


Figure 10. Correlations between performance metrics (rows) and the measured team features (columns). The correlation scores range from -1 (strongly negatively associated) to +1 (strongly positively associated). The correlation value is indicated when the correlation is significant at the $p=0.1$ level.

The features are divided into 4 types, relating to team composition (size, background diversity, prior experience with citizen science projects), activity on Slack (number of posts, communication with organisers, closeness centrality in the communication network), collaborations using CoSo (intra-team work interactions, amount and diversity of the advice-seeking ties), and activity reports using CoSo (amount and span of activities performed, regularity of work behaviour). The outcomes can be divided into 2 categories: the quality of the project (final stage progressed, relevance, novelty, ...), as judged by experts, and the quality of the process (engagement, deliverables), as judged by the organising team.

We note that there is an overall trend, where the quality of the projects is associated with team composition and Slack activity measures, while the quality of the process is associated with CoSo collaborations and activity measures. More precisely, for the team composition, we find that team size is associated with the use of at least a tool in the Citizen Science Solution Kit (mean_crowdsourcing), highlighting the need for a minimal human power available for achieving this task. The diversity of backgrounds in the team is associated with the novelty of the project, supporting findings that interdisciplinarity begets innovative work (Singh et al., 2021). Prior experience with citizen science is important for the relevance and

novelty of the project. Finally, we find that teams that progressed to later stages were composed of members who did not know each other before, which is in support of a potential role played by the algorithm used to form diverse and complementary teams.

In the case of Slack, we find that the overall activity, and in particular the number of interactions with the organising team, is associated both with stage progress as well as the relevance of the project. The team closeness centrality in the Slack interaction network, which quantifies its average short paths to other Slack users and therefore indicates a possibility for information aggregation, is also associated with performance (stage progressed), novelty and to have data potentially relevant to NSOs. Overall, these results highlight the role of mentoring for supporting team progress and focusing their projects.

For CoSo data, we find contrasting results. While the number of advice-seeking links to other participants is associated with project relevance (reflecting the findings from Slack data), we find that the derived features are indicative of performance related to *process* outcomes, as judged by the organising team. As such, teamwork and effort are associated with team commitment and their ability to produce qualitative deliverables and a successful final pitch. They are quantified both by the number of within-team collaborations and the number, diversity and regularity of activities performed, as well as their ability to integrate diverse advice, as measured by (lower) Burt constraint (Burt, 2004) in the advice-seeking network (the fact that a team's neighbours tend to represent different unconnected parts of the network). Teamwork and effort are also associated with the total score used to choose the top 5 teams progressing to the Accelerate phase. These results might indicate that the organising team, who was responsible to judge these criteria, was particularly sensitive to the ability of teams to engage and collaborate throughout the cycle, an information that was not readily available to most experts.

We note that none of the independent variables was significantly associated with the feasibility score, which might be either the result of a low statistical power, or indicating that the early stage of advancement at Evaluate does not allow for this outcome to be precisely measured. Similarly, the presentation score is not predicted by the measured features, which might indicate that it relies on other non-measured characteristics having to do more with participant personality traits.

5. Discussion and perspectives

In this report, we show how different data sources on teamwork, effort, communication and collaborations inform on various measures of performance of their project. The early stage of the projects did not allow us to measure their eventual level of robustness, the quality of the knowledge and solutions they produce or their interaction with a community of citizen scientists. Instead, we discuss below some insights gained from the second GEAR cycle on the performance of these projects at early stages, and how they can be integrated for the forward-thinking of GEAR 3.

First, we noted in GEAR 1 that Slack offers a limited capacity to illustrate collaborative behaviours, especially when there is low activity. Here, we used the CoSo app to collect self-reported information on collaborations, allowing us to delve further into team processes at play during GEAR 2. We also used a certain number of google forms, allowing us to easily adapt them last minute, and share their results. However, from a participant perspective, it is yet another tool to use. As a solution, we consider simplifying the survey design aspect of CoSo to integrate all forms within this tool. Furthermore, while engagement of participants to surveys was good, it could still be improved. A solution would be to allocate time for participants to answer the CoSo survey by including a 10 minutes slot in the program each week to do it with participants.

Second, we considered various metrics of performance related to both the outcome and the process. One outcome that we did not consider but could be of interest would be the learning of participants during the cycle: a measure of educational performance of Crowd4SDG. The ability to produce a most advanced prototype could also be considered, incentivizing for having a concrete first outcome. Moreover, the current stage advancement is based on a *numerus clausus* requirement, with only 5 teams selected for the CBIW Accelerate program. Since the workshop is online and that we don't have "hosting capacity constraints", it could be interesting to switch towards an absolute score based method and invite all projects passing a threshold score to integrate CBIW.

Finally, a clear limitation of this work is the low sample size used for association analyses. While trends allow us to gain a general understanding of how some indicators can be monitored to follow teamwork and progress, the level of significance ($p=0.1$) does not allow for clear-cut generalizable results. The comparison (and potential aggregation) with GEAR 3 will allow us to test whether these trends can be confirmed and strengthened.

6. Conclusion

Work Package 4 aims to develop and monitor new metrics and develop statistical models of team engagement and collaboration that contribute to the many-faceted outcomes of the citizen science projects developed within the Crowd4SDG consortium. In this report, we presented a data-driven approach to describe the GEAR cycle 2 through team composition, activity and interaction dynamics. In particular, we leveraged the CoSo platform for collecting self-reported data on collaborations and task allocation structure of participating teams. These findings serve as a basis for i) exhibiting the potential of using digital traces to derive measures related to team process, ii) highlighting perspectives for monitoring metrics in the next GEAR cycle. Hence, this work results from the efforts put into the Task 4.2 and 4.3 and is complemented by the Deliverable 4.2.

While our initial hope was to follow the *in-situ* dynamics of teams participating in the program, the unexpected shift to a fully online program led us to shift gears, focusing on digital traces from the team coordination tools, as well as building a smartphone application to facilitate the reporting of collaborative activities. By leveraging a single sign-in system and the ability to run all future surveys in one place using the CoSo app, we were able to collect comprehensive profile information about participants along with temporal insights on their collaborative activities, avoiding extensive manual curation from the organising team and ensuring analysis-ready data. Interestingly, we show how the CoSo allows us to monitor features that are associated with aspects of performance related to the *process* and not just the final outcome. Moreover, we show that beyond background diversity (important for project novelty), the diversity of advice-seeking ties (a network diversity metric) matters for the final project performance.

Overall, the quantitative description of the diversity of participants and teams, and the monitoring of their activity via the use of digital traces and self-reports provide the ground for the Crowd4SDG consortium to take evidence-based decisions for the next GEAR cycle and frame the improvement of citizen science skills and outcomes of the developed projects. The self-reported and surveyed data offer an opportunity to operationalise metrics and descriptors underlying the quality and novelty of citizen science projects. Our contribution extends beyond the Crowd4SDG project to the general evaluation of citizen science by informing project leaders, citizen scientists and decision makers on what can be assessed online to perform high-quality citizen science based on the criteria provided in D4.2 and operationalized in this report.

Hence, we make the following recommendations to the GEAR cycle 3 organising team based on our findings: i) promote diverse backgrounds within teams, and allow for larger team sizes ii) encourage teams to have diversity in terms of age and education levels, as these are markers of more successful teams, iii) foster regular engagement between participants, mentors and members of the organising team, iv) encourage collaboration and advice seeking across diverse teams, v) advise teams to work regularly on their project and allocate sufficient time to perform the wide range of activities needed for project completion and success.

7. References

Burt, R.S., 2004. Structural Holes and Good Ideas. *Am. J. Sociol.* 110, 349–399. <https://doi.org/10.1086/421787>

Poquet, O., Tupikina, L., Santolini, M., 2020. Are forum networks social networks? a methodological perspective, in: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, LAK '20*. Association for Computing Machinery, New York, NY, USA, pp. 366–375. <https://doi.org/10.1145/3375462.3375531>

Singh, C., Barme, E., Ward, R., Tupikina, L., Santolini, M., 2021. Quantifying the rise and fall of scientific fields. *ArXiv210703749 Phys.*

Tackx, R., Blondel, L., Santolini, M., 2021. Quantified Us: a group-in-the-loop approach to team network reconstruction, in: *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, UbiComp '21*. Association for Computing Machinery, New York, NY, USA, pp. 502–507. <https://doi.org/10.1145/3460418.3479363>



Annex 1: list of abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CBI	Challenge-based Innovation (in-person coaching)
CBix	Challenge-based Innovation (remote location)
CoSo	Collaborative Sonar
CS	Citizen Science
CSSK	Citizen Science Solution Kit
GEAR	Gather, Evaluate, Accelerate, Refine
NSO	National Statistical Office
O17	Open Seventeen Challenge (online coaching)
SDG	Sustainable Development Goal
WP	Work Package



Annex 2: Surveys

[Registration Form](#)

Evaluate weekly 1, 2, 3 and Accelerate

Which task(s) did you perform since the previous block?

- Brainstorming / ideation
- Planning tasks
- Developing protocols / methods
- Reading papers / documents
- Attending meetups with other teams
- Collaborating with other teams
- Interpreting results
- Preparing other documents/materials to share outside the team
- Preparing graphics / videos
- Writing / preparing a pitch
- Project administration
- Software development
- Hardware development
- Data analysis / modeling / visualization
- Meeting with people affected by the problem you are trying to solve
- Meeting with actors of the field of your solution
- Team Building activities
- Documentation of your project
- Use of CS project builder
- Use of CS logger
- Use of Decidim
- Use of Visual Cit
- Use of SDG in progress

0%

Who did you work with since the previous block?

Select collaborators within your team or from other teams. You may need to scroll down to see all of them.

Participants



Who did you seek advice from during the previous block?

Select the persons by name from the list below. You may need to scroll down to see all of them.

Search the participants

Participants

How helpful did you find the session on [Name of the session]?

- Did not attend & did not watch the Zoom recording
- Very unhelpful
- Unhelpful
- Neither unhelpful nor helpful
- Helpful
- Very helpful



Back

Next

What went especially well, or what could have gone better and how? Any other comments welcome too.

Text input area for comments.



Back

Next



[Final form - Evaluate](#)

Accelerate initial Form and bloc 1

Teamwork / collaboration
I am able to efficiently collaborate in a team.
Please rate accordingly.

- Strongly disagree
- Disagree
- Neither disagree or agree
- Agree
- Strongly agree

0%

Back Next

Communication
I am able to communicate my ideas or projects clearly and concisely.
Please rate accordingly.

- Strongly disagree
- Disagree
- Neither disagree or agree
- Agree
- Strongly agree

13%

Back Next



Problem solving
I am able to find solutions to challenges my problem is facing through brainstorming, prototyping, and / or engaging with stakeholders.
Please rate accordingly.

- Strongly disagree
- Disagree
- Neither disagree or agree
- Agree
- Strongly agree

25%

Back Next

Leadership / self-confidence
I feel confident in being able to launch my project.
Please rate accordingly.

- Strongly disagree
- Disagree
- Neither disagree or agree
- Agree
- Strongly agree

38%

Back Next

How helpful did you find the session on [Name of the session]?

- Did not attend & did not watch the Zoom recording
- Very unhelpful
- Unhelpful
- Neither unhelpful nor helpful
- Helpful
- Very helpful

38%

Back Next



What went especially well, or what could have gone better and how? Any other comments welcome too.

88%

Back

Next

[Accelerate Final Form](#)